Genetic and expression analysis of cattle identifies candidate genes in pathways responding to Trypanosoma congolense infection

Genetic and expression analysis of cattle identifies candidate genes in pathways responding to *Trypanosoma congolense* infection

Harry Noyes[a,1], Andy Brass[b,1], Isaiah Obara[c], Susan Anderson[d], Alan L. Archibald[d], Dan G. Bradley[e], Paul Fisher[b], Abigail Freeman[e], John Gibson[c,2], Michael Gicheru[f], Laurence Hall[d], Olivier Hanotte[c,3], Helen Hulme[b], Declan McKeever[g], Caitriona Murray[e], Sung Jung Oh[h], Catriona Tate[a,b], Ken Smith[g], Miika Tapio[c,4], John Wambugu[b], Diana J. Williams[i], Morris Agaba[c], and Stephen J. Kemp[a,c,5]

[a]School of Biological Sciences, University of Liverpool, Liverpool L69 7ZB, United Kingdom; [c]International Livestock Research Institute, Nairobi, 00100, Kenya; [e]Molecular Population Genetics, Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland; [d]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, Scotland, United Kingdom; [b]School of Computer Science/Faculty of Life Sciences, University of Manchester, Manchester M13 9PL, United Kingdom; [i]Department of Veterinary Pathology, University of Liverpool, Liverpool L69 7ZJ, United Kingdom; [g]Pathology and Infectious Diseases, Royal Veterinary College, Hatfield AL9 7TA, United Kingdom; [h]National Institute of Animal Science, Rural Development Association, Suwon 441-706, Korea; and [f]Department of Zoological Sciences, Kenyatta University, 00100, Nairobi, 00100, Kenya

Edited* by William C. Clark, Harvard University, Cambridge, MA, and approved April 21, 2011 (received for review September 21, 2010)

African bovine trypanosomiasis caused by *Trypanosoma* sp., is a major constraint on cattle productivity in sub-Saharan Africa. Some African *Bos taurus* breeds are highly tolerant of infection, but the potentially more productive *Bos indicus* zebu breeds are much more susceptible. Zebu cattle are well adapted for plowing and haulage, and increasing their tolerance of trypanosomiasis could have a major impact on crop cultivation as well as dairy and beef production. We used three strategies to obtain short lists of candidate genes within QTL that were previously shown to regulate response to infection. We analyzed the transcriptomes of trypanotolerant N'Dama and susceptible Boran cattle after infection with *Trypanosoma congolense*. We sequenced EST libraries from these two breeds to identify polymorphisms that might underlie previously identified quantitative trait loci (QTL), and we assessed QTL regions and candidate loci for evidence of selective sweeps. The scan of the EST sequences identified a previously undescribed polymorphism in ARHGAP15 in the Bta2 trypanotolerance QTL. The polymorphism affects gene function in vitro and could contribute to the observed differences in expression of the MAPK pathway in vivo. The expression data showed that TLR and MAPK pathways responded to infection, and the former contained TICAM1, which is within a QTL on Bta7. Genetic analyses showed that selective sweeps had occurred at TICAM1 and ARHGAP15 loci in African taurine cattle, making them strong candidates for the genes underlying the QTL. Candidate QTL genes were identified in other QTL by their expression profile and the pathways in which they participate.

nagana | positional cloning | sustainable agriculture | trypanosomosis

African trypanosomes are extracellular protozoan parasites that cause severe diseases in humans and livestock, usually with fatal consequences unless treated. Though *Trypanosoma brucei* spp. cause significant morbidity and mortality in humans, infections with *Trypanosoma congolense* and *Trypanosoma vivax* are among the most significant constraints on cattle production in Africa, causing major economic losses, which in turn have serious consequences for human health and welfare (1). Anemia is the most prominent and consistent clinical sign of infection and is the main indicator for treatment rather than parasitemia, which is highly variable, particularly in indicine cattle.

Some African *Bos taurus* cattle breeds, such as N'Dama, are tolerant of infection with *T. congolense*, remaining apparently healthy despite the presence of parasites. This capacity to remain productive while harboring potentially lethal trypanosome infections is known as trypanotolerance (2) and is an evolutionary adaptation between the host and pathogen (3). N'Dama cattle are native to West Africa, where they were introduced about 3,000 y ago (4), whereas the introduction of zebu (*Bos indicus*) cattle to the region is recent, probably only intensifying after a rinderpest epidemic just over a century ago (5).

Although N'Dama can tolerate moderate levels of trypanosome challenge, their temperament and relatively small size makes them unsuited for draft purposes. The use of cattle for pulling plows and carts can dramatically increase crop production, as well as provide, milk, meat, and manure. This loss of draft power means that many poor African farmers are dependent on the heavy labor of hand tillage, and suffer from loss of meat and milk production. Development of breeds of trypanotolerant cattle that are better adapted to the diverse needs of African farmers could substantially improve productivity in the whole agricultural system.

We are seeking to identify the genetic determinants of trypanotolerance in N'Dama with a view to introgressing this trait into zebu breeds. Ten major quantitative trait loci (QTL) that control trypanotolerance in cattle have been genetically mapped by our group using an F2 cross between Boran (zebu) and N'Dama cattle (6). However, the positions of the loci have been resolved only to within ±20 cm and lie in gene-rich regions that lack obvious functional candidates, or contain numerous genes with similar potential. In the present study we have used transcriptome analyses to identify pathways that are responding to infection in the liver, spleen, and precrural lymph nodes. We have combined these data with three different genetic analyses to identify mutations and signatures of selective sweeps in genes in those pathways that might

Author contributions: H.N., A.B., A.L.A., D.G.B., P.F., J.G., M.G., O.H., D.M., S.J.O., M.T., D.J.W., M.A., and S.J.K. designed research; I.O., S.A., A.F., L.H., C.M., K.S., M.T., J.W., and M.A. performed research; A.B. and P.F. contributed new reagents/analytic tools; H.N., A.B., I.O., D.G.B., P.F., A.F., H.H., C.M., C.T., K.S., M.T., and M.A. analyzed data; and H.N. and S.J.K. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: Array data were deposited in ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) under accession no. E-MEXP-1778. SNP were deposited in dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/) under ssIDs 196313094–196313199 and 196003459–196003467. EST sequences are available from EMBL (http://www.ebi.ac.uk/embl/) under accession nos. AM003907–AM039430.

[1]H.N. and A.B. contributed equally to this work.

[2]Present address: The Institute for Genetics and Bioinformatics, University of New England, Armidale NSW 2351, Australia.

[3]Present address: Department of Biology, University of Nottingham, NG7 2RD United Kingdom.

[4]Present address: MTT Agrifood Research, FI-31600, Jokioinen, Finland.

[5]To whom correspondence should be addressed. E-mail: kempsj@liv.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1013486108/-/DCSupplemental.

regulate the differences in response to infection. This data has been combined with prior knowledge from our linkage mapping studies and genetic and expression analyses of O'Gorman et al. (7) and Gautier et al. (8) to identify candidate genes in the five QTL of largest effect that have also been mapped to sufficiently small chromosomal regions to make the identification of candidate genes practicable.

Results

Candidate Quantitative Trait Genes. Hanotte et al. (6) localized QTL for parasitemia, body weight, and packed cell volume after *T. congolense* infection in a Boran × N'Dama F2 cross. We used four strategies to identify candidate quantitative trait genes (QTG) within the large QTL regions. First, we screened an EST library for nonsynonymous SNP in genes within the QTL; second, we used gene expression data from spleen, liver, and lymph nodes at nine time points to identify pathways that responded to infection and genes within those pathways that were in QTL; third, we undertook a genetic screen of 17 loci to discover evidence of selection; and fourth, we screened two QTL using published SNP data from the bovine HapMap for signatures of selection.

EST Library Screen Identifies an ARHGAP15$^{282H\rightarrow P}$ Mutation in the Bta2 QTL. We obtained lists of genes that are in QTL regions and then screened EST libraries made from four tissues from N'Dama and Boran for SNP within those genes. This procedure identified two synonymous SNP and one nonsynonymous SNP (nsSNP) in the *ARHGAP15* (Rho GTPase-activating protein 15) gene. No other nsSNPs were discovered in the other 30 genes in the Bta2 QTL. The nsSNP causes an H→P mutation at ARHGAP15^{282} and was predicted to be "probably damaging" by polymorphism phenotyping (PolyPhen) (9). Boran cattle carry the ancestral H allele at high frequency (0.71, $n = 28$), which is conserved in humans, pigs, chickens, and salmon, the *P* allele appears fixed in N'Dama from Guinea ($n = 30$). We have not found any records of the P allele in other species. ARHGAP15 is a RAC binding protein, and the mutation is at the proximal end of the RAC binding domain (PS50238) of the protein (10), suggesting a possible mechanism for any effect of the mutation on function. RAC1 is a GTPase that is active when binding GTP and inactive when binding GDP; the RhoGAP domain of ARHGAP15 accelerates the intrinsic rate of hydrolysis of GTP to GDP, thereby inactivating RAC1 (10). To determine whether the ARHGAP15$^{282H\rightarrow P}$ polymorphism might cause a difference in the rate of GTP hydrolysis by RAC1, the two alleles were expressed in vitro, and the rate of GTP hydrolysis was measured by following the change in absorbance at 360 nm caused by an increase in γP_i abundance (Fig. 1). Both alleles significantly increased the rate of GTP hydrolysis by RAC1 over the rate of the hydrolysis by RAC1 alone; however, the ARHGAP15^{282P} allele increased the rate of hydrolysis by 21% more than the ARHGAP15^{282H} allele over three replicate protein extractions and activity assays ($F_{3,\ 25.0.9} = 56.40$, $P < 0.0001$).

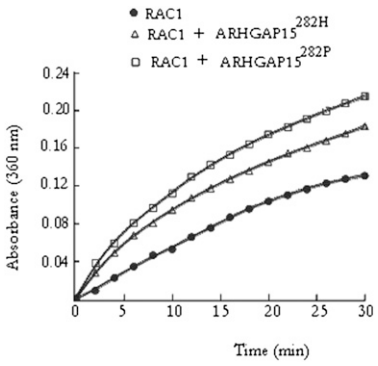

Fig. 1. Absorbance profile for the intrinsic RAC1, ARHGAP15^{282H}, and ARHGAP15^{282P} stimulated GTPase reactions and their fitted profiles. The 95% confidence limits were very small and are shown as a fine dotted line close to the fitted line.

Response of the Transcriptomes of N'Dama and Boran to Infection. Twenty-five animals of each breed were included in the expression analysis experiment. Five animals of each breed were killed before infection, and on days 21 and 35 postinfection (pi) for collection of liver, spleen, and precrural lymph node samples. Needle liver biopsies were taken from five animals of each breed at six additional time points (days 12, 15, 18, 26, 29, and 32 pi). Ten animals of each breed were only used for needle biopsies.

Analysis of the expression data focused on identification of KEGG pathways (11) that either responded to infection or differed between breeds. To identify candidate genes in each QTL we used a Taverna workflow to obtain a list of all genes that were in the KEGG pathways that contained an excess of genes ($P < 0.05$) that either changed over time in response to infection or were differentially expressed between breeds. The workflow then obtained a list of all genes within the QTL region from Ensembl (Assembly Btau3) irrespective of whether they were differentially expressed (12–14). These two lists were then scanned for genes that were in a responding pathway and in a QTL region; such genes were considered as candidate QTG. The most significant ($P < 0.05$) pathways and genes that were identified in this way for the five QTL of largest effect are shown in Table 1.

TICAM1 (TRIF), IKBKB (IKKB), ECSIT, VAV1, and *CASP8* were under QTL and in pathways that responded to infection (Table 1). The first four genes are important in the NFKB pathway; ECSIT bridges TRAF6 to MAP3K1, and the MAPK pathway and TICAM1 participates in the MYD88-independent alternative pathway (15). We found evidence that *TICAM1* has been under positive selection in West African taurine cattle (see below), suggesting that this gene may be an important regulator of response to infection by altering the balance between the MYD88 classical TLR response and the MYD88-independent alternative response (16). Differences in TICAM1 activity could be causing the differences we see in the expression of some or all of the other genes in this pathway.

VAV1 is an agonist for RAC1and is most highly expressed in CD14$^+$ monocytes (17); in contrast, ARHGAP15 is an antagonist for RAC1. *VAV1* was twofold more highly expressed in the liver of Boran than N'Dama, and if regulated by SNP within or close to the gene, then *VAV1* is a candidate QTL gene; however, it may also be indirectly regulated as a consequence of the difference in activity of ARHGAP15. *VAV1* was resequenced in Boran and N'Dama, but no structural differences were observed.

Responding KEGG Pathways. We also obtained lists of responding pathways, irrespective of whether they contained genes within QTL, to obtain a more comprehensive view of the response to infection. Five immune-related categories were strong responders in the liver of both breeds: natural killer (NK) cell-mediated cytotoxicity, T-cell receptor signaling pathway, B-cell receptor signaling pathway, Fc epsilon RI signaling pathway, and leukocyte transendothelial migration (Dataset S1). Of the immune-related categories, NK cell-mediated cytotoxicity was by far the most significantly overrepresented class of genes, and showed two- to fourfold up-regulation of all genes that were differentially expressed (Fig. 2). This up-regulation is most likely a marker for increased NK cell abundance rather than increased gene expression by resident NK cells, and corresponds to a lymphocyte-rich inflammatory reaction that was apparent on histological examination of the liver. The MAPK pathway, which is highly expressed in NK cells, was particularly strongly up-regulated and contains VAV1 and RAC1. MAPK pathway gene expression increased strongly and consistently in both breeds, suggesting that it has an important role in the response to infection. ARHGAP15 acts as a negative regulator of RAC1, which is a key mediator in this pathway. The combined effect of higher expression of *VAV1*,

SUSTAINABILITY SCIENCE

Table 1. Pathways that responded to infection in liver, spleen, or lymph node ($P < 0.05$ for at least one condition), and genes within each pathway also within each of the QTL

		QTL location and trait affected (4)				
KEGG pathway ID	KEGG pathway name	BTA2 anemia	BTA4 parasitemia	BTA7 parasitemia and anemia	BTA16 anemia	BTA27 anemia
04660	T-cell receptor signaling pathway	*CD28* *CTLA4* *ICOS*		*VAV1*		*IKBKB*
04620	Toll-like receptor signaling pathway	*CASP8*		*IRAK1* *TRIF*		*IKBKB*
04670	Leukocyte transendothelial migration			*VAV1*		*CLDN23*
04514	Cell adhesion molecules (CAMs)	*CD28* *CTLA4* *NLGN1* *ICOS*		*MADCAM1* *ICAM1* *ICAM3*		*CLDN23*
04640	Hematopoietic cell lineage			*EPOR* *FCER2*		
04210	Apoptosis	*CASP8*		*IRAK1* *PRKACA*	*CAPN2*	*IKBKB*
00561	Glycerolipid metabolism	*LCT*	*DGKI*			*AGPAT6*
00564	Glycerophospholipid		*DGKI*	*ARD1A*		*AGPAT6*
04110	Cell cycle	*ORC4L*, *MCM6*		*CDKN2D*		
04010	MAPK signaling pathway	*CASP8*	*CASP2*	*PRKACA* *ECSIT*	*DUSP10*	*BRAF* *DUSP4* *IKBKB* *FGF20*
03010	Ribosome			*RPL36* *RPS28*		
04810	Regulation of actin cytoskeleton	*FN1* *PIP5K3*	*CHRM2*	*VAV1*		*BRAF* *FGF20*
04540	Gap junction			*TUBB4* *PRKACA*	*GNAQ*	
04510	Focal adhesion	*FN1*	*ZYX*	*VAV1* *COL5A3*	*CAPN2*	*BRAF*
00480	Glutathione metabolism	*IDH1*	*GSTK1*			
04520	Adherens junction			*INSR*		*FGFR1*
01510	Neurodegenerative disorders	*CASP8*				
04720	Long-term potentiation			*PRKACA*	*GNAQ*	*BRAF*
04310	Wnt signaling pathway			*PRKACA*		*DKK4* *SFRP1*
04360	Axon guidance	*EFNB1*	*EPHA1* *EPHB6*			*UNC5D*
04020	Calcium signaling pathway		*CHRM2*	*PTGER1* *PRKACA*	*GNAQ* *GNA14* *ITPKB*	*ADRB3* *VDAC3*

The genes within QTL that are listed were not necessarily differentially expressed, because the QTL effect on the pathway could be a consequence of a structural polymorphism that does not affect expression. Note that *ARHGAP15* does not appear in this list only because KEGG have yet to include it in the MAPK pathway. A list of complete gene names is available in Table S1. [Reproduced with permission from ref. 31 (Copyright 2008, Karger, Basel).]

a RAC1 agonist, in Boran and higher activity of *ARHGAP15*, a RAC1 anatagonist, in N'Dama might lead to a stronger MAPK-mediated inflammatory response in Boran than N'Dama.

Pathways Associated with Differentially Expressed Genes. The above analysis was undertaken on genes that differed between breeds as well as those that that did not differ between breeds but changed in expression in response to infection. The genes that differed in expression between the breeds were tested separately to identify the KEGG pathways that differed in response. Cytokine-cytokine receptor interaction, MAPK signaling pathway, and neurodegenerative disease were the only pathways containing an excess of differentially expressed genes and were only differentially expressed in the lymph nodes (Dataset S1). The presence of the MAPK pathway in the list of pathways associated with differentially expressed genes as well as the pathways associated with responding genes emphasizes its potential importance and is consistent with observations from previous studies of PBMC from the same animals (7).

Signatures of Selection Around TICAM1 and ARHGAP15. The trypanosomiasis challenge of West Africa has undoubtedly been a source of selective pressure on cattle as they spread from the largely trypanosome-free Sahel into intense challenge in West Africa within the last 3,000 y (18). Trypanotolerant populations might be expected to show signatures of selection around genetic variants responsible for that trait. Therefore, we used the data from the Bovine SNP consortium to test the QTL that contained

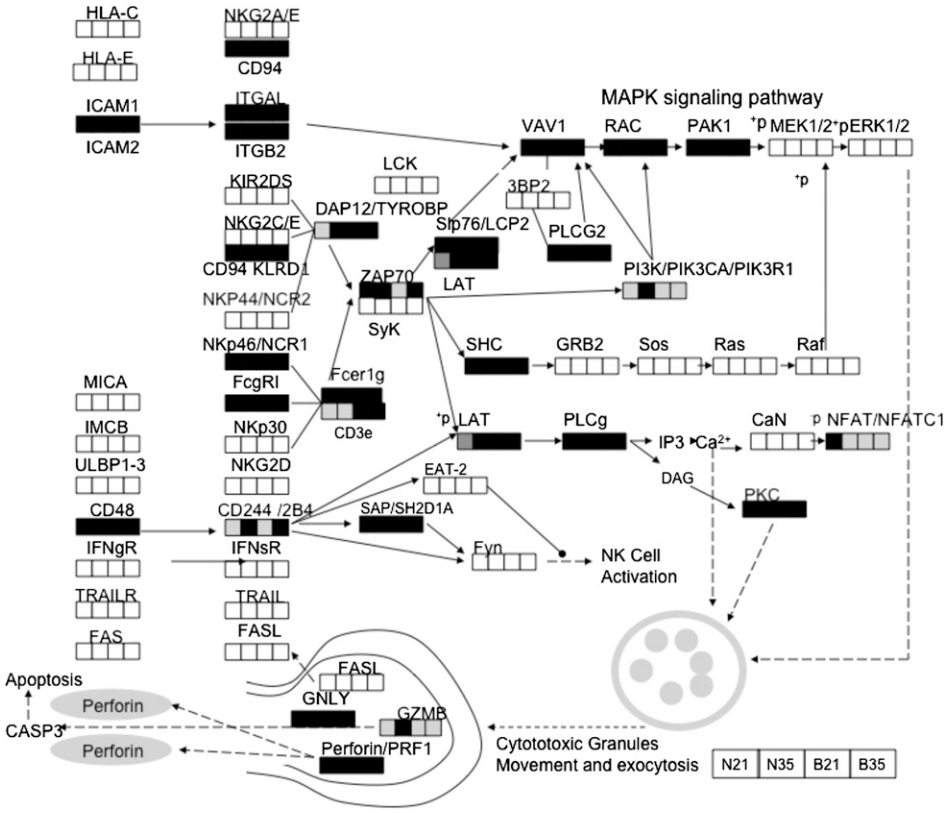

Fig. 2. KEGG NK cell pathway (hsa04650) in the liver showing up-regulation ($P < 0.05$, fold change > 2; black); no significant change (gray); not represented in array (white). No genes were down-regulated. Within each block, conditions are represented in the following order: (*i*) N'Dama change from day 0–21; (*ii*) N'Dama change from day 0–35; (*iii*) Boran change from day 0–21; (*iv*) Boran change from day 0–35. Multiple receptors are up-regulated, and the MAPK pathway from the receptors ITGAL/ITGB2 through VAV1, RAC, and PAK1 are all consistently up-regulated in both breeds.

ARHGAP15 and *TICAM1* for evidence of a selective sweep. The data included the East African Sheko and West African N'Dama, but not Boran. Sheko is commonly considered a taurine breed, but the HapMap data indicates that the majority of alleles are of zebu origin (19). We separately tested for a signature of selection at *TICAM1* in a panel of *B. indicus* and European and African *B. taurus* breeds.

Selective sweeps can influence wide areas and local evidence might sometimes originate from selection on linked loci. However, the cross-population extended haplotype homozygosity (XP-EHH) test (20) for extended chromosome segment variants that have reached high frequency suggested narrow selective sweeps in Sheko around *ARHGAP15*, but not around the *TICAM1* gene (Fig. 3). For the 0.5-megabase (Mb) segment with *ARHGAP15*, the XP-EHH score was in top 2.3% in the genome, but *TICAM1* was not an outlier, being only in the top 22%. The *ARHGAP15* sweep area included only one other gene, glycosyltransferase-like domain containing 1 (*GTDC1*), which shared the same signal of selective sweep. *GTDC1* is ubiquitously expressed and could conceivably affect the response to infection, but there was no other evidence to implicate it. Use of other reference breeds (Brahman, Gir, Holstein, and Jersey) resulted in similar signals for positive selection at these QTL, although the peak shapes demonstrated some variation. These results provide evidence for a rapid completed positive sweep at *ARHGAP15* in West African *B. taurus*.

We genotyped multiple animals of each breed to identify the frequencies of the resistant $ARHGAP15^{282P}$ allele and found Guinea N'Dama 1.0 ($n = 30$), Sheko 0.67 ($n = 18$), and Kenyan Boran 0.29 ($n = 28$). Although it is not easy to quantify the trypanotolerance of these animals, these frequencies do at least correlate with how these breeds are ranked for this trait, and in the case of Sheko and Boran are about twice the allele frequencies that would be expected from the percentage of taurine alleles in their genomes in the HapMap data (19).

We also investigated *TICAM1* for a signature of selection by resequencing the gene plus neighboring noncoding regions in a panel of 10 African *B. taurus*, 18 European *B. taurus*, and 11 Indian *B. indicus*. This gene exhibited outlying behavior with 17 segregating nsSNPs, two of which were scored as "probably damaging" by PolyPhen ($TICAM1^{268P\rightarrow L}$ and $TICAM1^{294P\rightarrow L}$) and are therefore likely to have an effect on function. Seven of the coding changes in *TICAM1* define two highly divergent segregating haplotypes. One of these predominates in the African sample and shows a pronounced excess of derived alleles, reflected in a highly significant value when Fay and Wu's *H* test of neutrality is applied

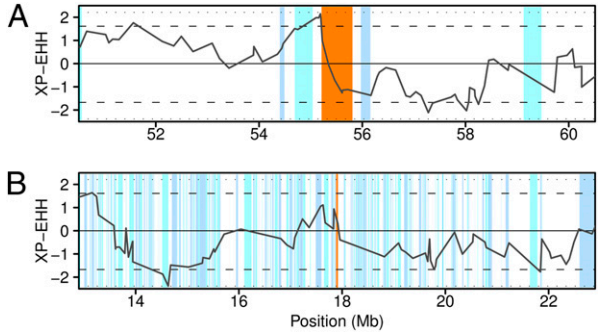

Fig. 3. Standardized cross-population extended haplotype homozygosities (XP-EHH) >10 Mb around *ARHGAP15* (*A*) and *TICAM1* (*B*) whose location in the center is marked with an orange column. Extreme positive values imply positive selective sweep in N'Dama, and negative values suggest sweep in Sheko. Genome-wide first and 99th percentiles across all XP-EHH scores are indicated with dotted line; fifth and 95th percentiles are indicated with hashed line. Alternating pale green and blue bars indicate other genes in the region (excluding hypothetical unrecognized genes). *ARHGAP15* is in a gene-poor region with a broad peak of positive selection, whereas *TICAM1* is in a gene-rich region with a narrow peak of selection that does not reach the 95% confidence threshold. At *ARHGAP15* and *TICAM1*, the median intermarker intervals were 107 and 87 kb, respectively.

to the African population ($H = -19.17$, $P = 0.001$) but not to the European or Indian populations. The locus shows a high level of linkage disequilibrium throughout the 4 kbp sequenced. Interestingly, each of these behaviors was extreme when viewed within the context of results for a total of 18 loci that were wholly or partially resequenced in similar continental cohorts (21, 22) (Table S2). These included *TNFAIP8L1*, a second immunogene also targeted from under the chromosome 7 QTL peak, which gave no outlying behavior, plus several other loci of immunological consequence (e.g., several interleukins and Toll-like receptors), some of which also give indications of adaptive history. These population genetic values at *TICAM1* provide evidence for a local selective sweep within African *B. taurus*.

Discussion and Conclusion

This combination of genetic, in vitro analysis and expression data has provided unique insights into the differences between susceptible Boran and tolerant N'Dama cattle in their response to infection with *T. congolense*. We see clear evidence of differential innate responses early in the infection, apparently mediated by sequence differences in previously identified QTL regions. The pathway analysis suggested an important role for NK cells in the response to infection in both breeds and differential expression of the MAPK pathway. We have also identified a copy number variation in the murine *Cd244* NK cell receptor gene as a candidate QTL gene for response to *T. congolense* infection in mice, suggesting that NK cells may play an important role in the response to infection in mice as well as cattle (23). A map kinase (*MAPK14*) was differentially expressed in two-color spotted array study of PBMC obtained from the same animals at the same time points and may be indicative of the MAPK pathway being differentially expressed in PBMC as well (7).

There is compelling evidence that a nonsynonymous sequence difference in the *ARHGAP15* gene is playing an important role in the response and that it is the sequence variant that gives rise to the QTL on Bta2 in the mapping population. The putatively tolerant allele has altered behavior in vitro relative to the typical *B. taurus* allele, and it would be expected to inhibit RAC1 activity in the MAPK pathway, which could lead to the observed differences in expression or amplify downstream expression differences caused by other factors. There is also evidence for a selective sweep at this locus that includes just one other gene.

Similarly, KEGG pathway analysis of expression data identified *TICAM1* in the Toll-like receptor pathway as a candidate gene on Bta7, and genotyping identified *TICAM1* as being under positive selection in African taurine cattle. *TICAM1* contained two nsSNP that would probably significantly affect function. An independent whole genome scan with 36,000 SNP for signatures of selection in West African cattle recently identified *TICAM1* as a major hub in a network of genes under selection (8). Therefore, one analysis presented here and an independent study both indicate that TICAM1 is under selection in at least some African cattle breeds, including N'Dama, although there was no evidence from the XPEHH test for a sweep at *TICAM1* in Sheko. Trypanosomiasis is probably the largest disease constraint on cattle in sub-Saharan Africa, and this fact alone makes it likely that it is trypanosomiasis that has been driving this selection. In addition, *TICAM1* is also under the peak of a QTL associated with lower parasitemia in N'Dama, and our systematic analysis of expression data flagged TICAM1 as being in a pathway responding to infection with *T. congolense*. These independent lines of evidence make it a strong candidate gene at this locus. In this case, two nonsynonymous differences between the resistant and tolerant alleles are predicted to significantly change the gene's characteristics.

Trypanotolerance is an economically important trait of great biological interest, and an understanding of its mechanisms has the potential to transform cattle-keeping in tsetse-affected areas of Africa. However, trypanotolerance is a complex quantitative trait, with strong interactions with environmental and other disease factors (24). Identification of quantitative trait genes is particularly difficult because trypanosomes are free-living bloodstream parasites that interact with all tissues, causing a generalized response with cachexia and anemia but no pathognomonic signs that could be related to particular pathways or cell types. We have used microarrays to give us objective signatures of infection, and we have used these to identify the major pathways that respond to infection. We have investigated two genes that are in these pathways and under QTL by additional methods and show that both *ARHGAP15* and *TICAM1* appear to have functional polymorphisms that could affect the response to infection and are associated with signatures of selection. This is persuasive evidence that these genes are QTL genes, but definitive proof of their role in determining trypanotolerance status must await either a transgenic approach or a breeding program, which exploits them and in the process identifies recombinants. Such a breeding program could also produce cattle that could significantly increase the productivity of arable as well as livestock-based farming systems by making draft power available to all African farmers.

Methods

Animals, Design, Infection, and Sample Collection. All animal experiments were conducted in compliance with the International Livestock Research Institute (ILRI) ethical review process. Twenty-five susceptible (Boran) and 25 trypanotolerant (N'Dama) cattle were selected from the herd at the ILRI Kapiti Plains ranch, which is free from tsetse flies and trypanosomiasis. All animals were screened and confirmed to be negative for infection with tick-borne parasites before being transferred to the ILRI research facility at Kabete. On arrival, the cattle were treated with an anthelmintic (albendazole), sprayed with acaricide, and quarantined for 2 wk. They were allowed to acclimatize for at least 6 wk and maintained on a diet of hay supplemented with minerals and concentrates.

Cattle were infected by tsetse-mediated inoculation of trypanosomes with *T. congolense* IL1180 as previously described (25), and infection was confirmed by microscopy. Tissue specimens were collected by biopsy (liver) or at postmortem (liver, spleen, and precrural lymph nodes). For biopsy collection, animals were sedated with 0.5–0.7 mL of 20% xylaxine, and locally anesthetized with 10–15 mL of lignocaine. Liver biopsy samples (75 mg) were collected using Quick-Core biopsy needles through an incision made between the 10th and 11th ribs.

All tissues were collected into cryotubes and preserved in liquid nitrogen. The entire experiment was performed in four phases over 4 mo. Total RNA was extracted from tissues pulverized under liquid nitrogen using the TRIzol method (Gibco). Total RNA was treated with DNaseI, purified with RNAeasy (Qiagen), and checked for RNA integrity number >6.3 using an Agilent Bioanalyzer 2000.

Samples of liver and spleen collected into cryotubes for histopathology were postfixed in neutral buffered formalin, sectioned at 6 μm, and stained with H&E for light microscopy.

Expression data were acquired on Affymetrix Bovine Genome arrays, which have probesets for 23,000 transcripts. Samples were hybridized in random order. A total of 5 μg of RNA preparation was labeled using the One-Cycle Eukaryotic Target labeling protocol (Affymetrix), hybridized to Bovine Genome array chips, and scanned with an Affymetrix Gene Chip 3000 scanner. Any sample that failed the labeling or hybridization data quality control using dChip (26) or principal component analysis using SVD in the MAXD package (27) was discarded and the procedure repeated on a new RNA preparation from the original tissue sample. A total of 160 RNA samples passed quality control and the data were normalized using robust multichip average. Additionally, sets of 25-mer probes for each gene represented on the array were identified using AffyProbeMiner (28) and normalized in the R environment using multi-mgMOS (29).

The DAVID database (30) was used to identify responding KEGG pathways (Dataset S1). A Taverna workflow (12–14) was used to identify KEGG pathways that responded to infection and contained genes in QTL regions as previously described (31).

Comparisons between breeds were based on the change in expression over time for each breed as calculated by Eq. 1. This strategy favors detection of KEGG categories that differ in response to infection rather than having baseline differences in expression. Lists of genes for which $|d| > 2$ was submitted to DAVID to identify KEGG categories that contained an excess of differentially expressed genes. The threshold of $|d| > 2$ corresponds to 2.0 SDs of d over the entire dataset (Dataset S1).

$$d_{2I} = (B_{21} - B_0) - (N_{21} - N_0) \quad [1]$$

In Eq. **1**, letters indicate breeds (N, N'Dama; B, Boran) and numbers indicate days pi. Each alphanumeric symbol indicates the mean absolute $\log_2$ signal from the five replicates of a given gene for a given condition. Hence $N_{21} - N_0$ is the log ratio of expression for a given gene at day 21 pi vs. that at day 0 pi in N'Dama.

Preparation of cDNA Libraries and EST Sequencing. RNA was prepared from liver, lymph node, spleen, and bone marrow of three Boran cattle (BW064 uninfected, BW062 slaughtered at day 15 pi, and BW063 day 50 pi) and three N'Dama cattle (ND159 uninfected, ND161 day 15 pi, and ND160 day 50 pi). Tissues were submitted to GATC Biotech Ltd. for RNA extraction, cDNA synthesis and normalization, and cloning. Libraries were constructed for each tissue, and tissue was identified by linker sequences. Five thousand clones were picked from each library at the Roslin Institute and submitted to the Sanger Centre for bidirectional sequencing. Vector-trimmed and quality-clipped sequences were deposited at European Molecular Biology Laboratory under accession nos. AM003907–AM039430.

ARHGAP15 Cloning and Hydrolysis Rate. *ARHGAP15* RT-PCR products from N'Dama and Boran liver cDNA were cloned into the pGEX-4T1 expression vector. GST-tagged protein was purified, and the rate of hydrolysis of GTP in the presence of the Boran or N'Dama alleles was followed by measurement of absorbance at 360 nm. A semiparametric generalized linear mixed model was used to model the time series of absorbances to allow for possible nonlinear and disparate trends for the intrinsic RAC1, ARHGAP15 (H), and ARHGAP15 (P) stimulated GTPase reactions and to test for differences in expected absorbance across the three experiments. The model assumed a normal error distribution and the identity link function. The trend model included the experiment as a fixed effect and a B-spline smoothing covariance structure with three equally spaced interior knots (32). The model was fitted by restricted log pseudolikelihood within a mixed-model framework in the SAS GLIMMIX procedure (33).

Cross-Population Extended Haplotype Homozygosity Test for Positive Sweeps in N'Dama. Published cattle HapMap SNP data for N'Dama and Sheko was used for exploring signatures of positive selection in N'Dama using the East African Sheko breed as the reference population. The cattle HapMap data consisted of 21,034 SNP after exclusion of poorly supported SNP, and included 25 N'Dama and 20 Sheko individuals (19). The median interval between markers was 65 kb, which increased to 85 kb if the three most densely genotyped chromosomes (6, 14, and 25) were excluded. At *ARHGAP15* and *TICAM1*, the median intervals were 107 and 87 kb, respectively. Haplotyping of 35-Mb segments was done using Shape-IT (34). Standardized XP-EHH values (20) were calculated as in Pickrell et al. (34). Exclusion of related animals had no noticeable effect. The data were specifically tested for positive sweeps at *ARHGAP15* and *TICAM1*, and a genome wide analysis was done only to provide the distribution of experimental *P* values; therefore, it was not necessary to correct *P* values for multiple testing.

Sequencing and Population Genetic Analysis of *TICAM1*. A panel of 10 African (Ndama, Somba, Lagune) *B. taurus*, 18 European (Aberdeen Angus, Friesian, Norwegian Red, German Black, Highland, Alentejana, Mertolenga, Romagnola, Sikias, Anatolian Black) *B. taurus*, and 11 Indian (Hariana, Tharparkar, Sahiwal, Ongole) *B. indicus* were sequenced as previously described (21). Plains bison (*Bison bison*) was used as an outgroup. The data were analyzed using Fay and Wu's *H* test (35).

Additional Files and Resources. All expression data are available from ArrayExpress under accession no. E-MEXP-1778. Plots of the expression from each probe set on the array are also available from the authors' Web site at http://www.genomics.liv.ac.uk/tryps/resources.html.

ACKNOWLEDGMENTS. We thank the staff of the International Livestock Research Institute Large Animal Unit for expert technical assistance, and the International Atomic Energy Agency for making the Sheko genotype data available. This work was supported by Wellcome Trust Grant GR066764MA (to S.J.K.), Science Foundation Ireland Grant 02-IN.1-B256 (to D.G.B., A.F., and C.M.), and Ondex Systems Approach to Biological Research (SABR) Grant BB/F006012/1 (to P.F.).

1. Kristjanson PM, Swallow BM, Rowlands GJ, Kruska RL, de Leeuw PN (1999) Measuring the costs of African animal trypanosomosis, the potential benefits of control and returns to research. *Agric Syst* 59:79–98.
2. Murray M, Trail JC, Davis CE, Black SJ (1984) Genetic resistance to African Trypanosomiasis. *J Infect Dis* 149:311–319.
3. Mattioli RC, Wilson RT (1996) Trypanosomes, tsetse and trypanotolerance: Coevolution in tropical Africa. *Parassitologia* 38:531–535.
4. Hassan F (2000) Climate and cattle in North Africa: A first approximation. *The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics and Ethnography*, eds Blech R, MacDonald K (UCL Press, London).
5. Epstein H (1971) *The Origin of the Domestic Animals of Africa* (Africana, New York).
6. Hanotte O, et al. (2003) Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proc Natl Acad Sci USA* 100:7443–7448.
7. O'Gorman GM, et al. (2009) Transcriptional profiling of cattle infected with *Trypanosoma congolense* highlights gene expression signatures underlying trypanotolerance and trypanosusceptibility. *BMC Genomics* 10:207.
8. Gautier M, et al. (2009) A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* 10:550.
9. Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: Server and survey. *Nucleic Acids Res* 30:3894–3900.
10. Seoh ML, Ng CH, Yong J, Lim L, Leung T (2003) ArhGAP15, a novel human RacGAP protein with GTPase binding property. *FEBS Lett* 539:131–137.
11. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32(Database issue):D277–D280.
12. Fisher P, et al. (2007) A systematic strategy for large-scale analysis of genotype phenotype correlations: Identification of candidate genes involved in African trypanosomiasis. *Nucleic Acids Res* 35:5625–5633.
13. Oinn T, et al. (2006) Taverna: Lessons in creating a workflow environment for the life sciences. *Concurr Comput* 18:1067–1100.
14. Hull D, et al. (2006) Taverna: A tool for building and running workflows of services. *Nucl Acids Res* 34(Suppl 2):W729–W732.
15. O'Neill LA, Bowie AG (2007) The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat Rev Immunol* 7:353–364.
16. Lee MS, Kim Y-J (2007) Signaling pathways downstream of pattern-recognition receptors and their cross talk. *Annu Rev Biochem* 76:447–480.
17. Su AI, et al. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 101:6062–6067.
18. Jousse H (2004) A new contribution to the history of pastoralism in West Africa. *J African Archaeology* 2(2):187–204.
19. Gibbs RA, et al.; Bovine HapMap Consortium (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324:528–532.
20. Sabeti PC, et al.; International HapMap Consortium (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449:913–918.
21. Murray C, Huerta-Sanchez E, Casey F, Bradley DG (2010) Cattle demographic history modelled from autosomal sequence variation. *Philos Trans R Soc Lond B Biol Sci* 365: 2531–2539.
22. Freeman AR, Lynn DJ, Murray C, Bradley DG (2008) Detecting the effects of selection at the population level in six bovine immune genes. *BMC Genet* 9:62.
23. Goodhead I, et al. (2010) A comprehensive genetic analysis of candidate genes regulating response to *Trypanosoma congolense* infection in mice. *PLoS Negl Trop Dis* 4:e880.
24. Courtin D, et al. (2008) Host genetics in African trypanosomiasis. *Infect Genet Evol* 8: 229–238.
25. Akol GW, Murray M (1983) *Trypanosoma congolense*: Susceptibility of cattle to cyclical challenge. *Exp Parasitol* 55:386–393.
26. Li C, Wong W (2003) DNA-Chip Analyzer (dChip). *The Analysis of Gene Expression Data: Methods and Software*, eds Parmigiani G, Garrett E, Irizarry R, Zeger S (Springer, New York), pp 120–141.
27. Hancock D, et al. (2005) maxdLoad2 and maxdBrowse: Standards-compliant tools for microarray experimental annotation, data management and dissemination. *BMC Bioinformatics* 6:264.
28. Liu H, et al. (2007) AffyProbeMiner: A web resource for computing or retrieving accurately redefined Affymetrix probe sets. *Bioinformatics* 23:2385–2390.
29. Liu X, Milo M, Lawrence ND, Rattray M (2005) A tractable probabilistic model for Affymetrix probe-level analysis across multiple chips. *Bioinformatics* 21:3637–3644.
30. Dennis G, Jr., et al. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4:3.
31. Rennie C, et al. (2008) A systematic, data-driven approach to the combined analysis of microarray and QTL data. *Developments in Biologicals* (Karger, Basel), Vol 132, pp 293–299.
32. de Boor C (2001) *A Practical Guide to Splines* (Springer, New York), Rev. Ed.
33. Ruppert D, Wand M, Carroll R (2003) *Semiparametric Regression* (Cambridge Univ Press, Cambridge, UK).
34. Pickrell JK, et al. (2009) Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* 19:826–837.
35. Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155: 1405–1413.

Supporting Information

Noyes et al. 10.1073/pnas.1013486108

Table S1. Abbreviations

Gene name	Description
ADRB3	Adrenergic, β-3-, receptor
AGPAT6	1-acylglycerol-3-phosphate O-acyltransferase 6 (lysophosphatidic acid acyltransferase, zeta)
ARHGAP15	Rho GTPase activating protein 15
BRAF	v-raf murine sarcoma viral oncogene homolog B1
CAPN2	Calpain 2, (m/II) large subunit
CASP2	Caspase 2, apoptosis-related cysteine peptidase
CASP8	Caspase 8, apoptosis-related cysteine peptidase
CD14	CD14 molecule
CD28	CD28 molecule
CDKN2D	Cyclin-dependent kinase inhibitor 2D (p19, inhibits CDK4)
CHRM2	Cholinergic receptor, muscarinic 2
CHRM2	Cholinergic receptor, muscarinic 2
CLDN23	Claudin 23
CTLA4	Cytotoxic t lymphocyte-associated protein 4
DGKI	Diacylglycerol kinase, iota
DKK4	Dickkopf homolog 4 (*xenopus laevis*)
DUSP10	Dual-specificity phosphatase 10
DUSP4	Dual-specificity phosphatase 4
ECSIT	ECSIT homolog (*Drosophila*)
ECSIT	ECSIT homolog (*Drosophila*)
EPHA1	EPH receptor A1
EPHB6	EPH receptor B6
EPOR	Erythropoietin receptor
FCER2	Fc fragment of IgE, low affinity II, receptor for (CD23)
FGF20	Fibroblast growth factor 20
FGFR1	Fibroblast growth factor receptor 1
GNA14	Guanine nucleotide binding protein (G protein), alpha 14
GNAQ	Guanine nucleotide binding protein (G protein), q polypeptide
GSTK1	GST kappa 1
ICAM1	Intercellular adhesion molecule 1
ICAM3	Intercellular adhesion molecule 3
ICOS	Inducible T-cell costimulator
IDH1	Isocitrate dehydrogenase 1 (nadp+), soluble
IKBKB	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta
INSR	Insulin receptor
TICAM1	Toll-like receptor adaptor molecule 1
ITPKB	Inositol 1,4,5-trisphosphate 3-kinase B
LCT	Lactase
MADCAM1	Mucosal vascular addressin cell adhesion molecule 1
MAPK14	Mitogen-activated protein kinase 14
MCM6	Minichromosome maintenance complex component 6
MAP3K1	Mitogen-activated protein kinase kinase kinase 1
MYD88	Myeloid differentiation primary response gene (88)
ORC4L	Origin recognition complex, subunit 4-like (yeast)
PIKFYVE	Phosphoinositide kinase, FYVE finger containing
PRKACA	Protein kinase, cAMP-dependent, catalytic, alpha
PTGER1	Prostaglandin E receptor 1 (subtype EP1), 42kDa
RAC1	Ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)
RPL36	Ribosomal protein l36
RPS28	Ribosomal protein s28
SFRP1	Secreted frizzled-related protein 1
TICAM1	Toll-like receptor adaptor molecule 1
TUBB4	microRNA 220b

Table S1. Cont.

Gene name	Description
UNC5D	unc-5 homolog D (*C. elegans*)
VAV1	vav 1 guanine nucleotide exchange factor
VAV1	vav 1 guanine nucleotide exchange factor
VDAC3	Voltage-dependent anion channel 3
ZYX	Zyxin

Table S2. Fay and Wu's *H* for 18 genes

Gene	Population	bp	*N*[a]	S[b]	SS[c]	SN[d]	Fay and Wu's *H*[e]
ART4	Africa	2,157	10	6	0	1	0.779
ART4	Europe	2,157	18	6	0	0	0.397
ART4	India	2,157	11	33	6	4	−2.338
	Total	2,157	39	38	6	5	
CD2	Africa	3,752	10	13	3	6	−1.2
CD2	Europe	3,752	18	30	6	10	−9.511*
CD2	India	3,752	11	29	6	10	−7.775
	Total	3,752	39	33	6	11	
FEZL	Africa	4,282	9	2	1	0	0.366
FEZL	Europe	4,282	18	8	0	1	0.898
FEZL	India	4,282	11	11	1	1	1.784
	Total	4,282	38	17	2	1	
IL2	Africa	3,461	10	1	0	0	−1.253*
IL2	Europe	3,461	18	12	0	0	−5.68*
IL2	India	3,461	11	7	1	0	0.156
	Total	3,461	39	14	1	0	
IL5	Africa	2,521	10	13	0	0	−12.547***
IL5	Europe	2,521	18	16	0	0	−12.400**
IL5	India	2,521	11	15	0	0	−3.151
	Total	2,521	39	0	0	0	
IL13	Africa	3,297	10	15	0	1	1.621
IL13	Europe	3,297	18	15	0	0	1.756
IL13	India	3,297	11	14	0	0	2.13
	Total	3,297	39	26	0	0	
MRPL30	Africa	1,220	6	2	0	0	0.017
MRPL30	Europe	1,220	18	9	1	1	−7.708*
MRPL30	India	1,220	11	7	1	0	3.117
	Total	1,220	35	11	1	1	
MRPS05	Africa	2,455	9	11	3	4	0.017
MRPS05	Europe	2,455	19	19	3	2	−1.477*
MRPS05	India	2,455	11	31	5	3	0.208
	Total	2,455	39	32	5	4	
MRPS14	Africa	628	11	2	0	0	−6.400**
MRPS14	Europe	628	20	1	0	0	0.2667
MRPS14	India	628	11	1	0	0	0.178
	Total	628	42	2	0	0	
PRF1	Africa	1,107	5	3	1	1	−4.8**
PRF1	Europe	1,107	5	1	0	1	0.267
PRF1	India	1,107	5	3	1	3	0.622
	Total	1,107	15	6	1	4	
TLR2	Africa	1,114	5	3	1	3	0.711
TLR2	Europe	1,114	5	8	1	6	−4.000*
TLR2	India	1,114	5	9	2	7	0.8
	Total	1,114	15	13	3	10	

Table S2. Cont.

Gene	Population	bp	N^a	S^b	SS^c	SN^d	Fay and Wu's H^e
TLR4	Africa	1,100	5	2	1	1	0.444
TLR4	Europe	1,100	5	2	1	1	−0.622
TLR4	India	1,100	5	15	12	3	−6.311*
	Total	1,100	15	16	12	4	
TLR5	Africa	412	5	0	0	0	
TLR5	Europe	412	5	0	0	0	
TLR5	India	412	5	1	1	0	0.178
	Total	412	15	2	1	0	
TLR7	Africa	826	5	0	0	0	
TLR7	Europe	826	5	0	0	0	
TLR7	India	826	5	0	0	0	
	Total	826	15	0	0	0	
TLR9	Africa	941	3	2	2	0	0
TLR9	Europe	941	4	1	1	0	−1.500*
TLR9	India	941	5	1	1	0	0.267
	Total	941	12	4	3	0	
TNFAIP8L1	Africa	2,163	10	2	0	0	0.337
TNFAIP8L1	Europe	2,163	18	10	0	1	−0.838
TNFAIP8L1	India	2,163	11	6	0	1	−0.987
	Total	2,163	39	12	0	1	
TICAM1	Africa	4,130	10	24	8	8	−19.179***
TICAM1	Europe	4,130	18	41	14	15	−7.137
TICAM1	India	4,130	11	22	6	8	−5.576
	Total	4,130	39	50	16	17	
TYROBP	Africa	3,966	10	8	0	2	−3.779*
TYROBP	Europe	3,966	18	20	1	1	−7.356**
TYROBP	India	3,966	11	12	0	0	1.887
	Total	3,966	39	27	1	2	

TICAM1 has the most extreme value of all genes tested.
[a]Number of individuals sequenced.
[b]Number of SNPs.
[c]Number of synonymous SNPs.
[d]Number of nonsynonymous SNPs.

Other Supporting Information Files

Dataset S1 (XLS)